AStA Adv Stat Anal (2017) 101:381–398
DOI 10.1007/s10182-017-0292-5

ORIGINAL PAPER

Species occupancy estimation and imperfect detection: shall surveys continue after the first detection?

Gurutzeta Guillera-Arroita[1] · José J. Lahoz-Monfort[1]

Received: 21 June 2015 / Accepted: 11 January 2016 / Published online: 17 February 2017

Abstract Species occupancy, the proportion of sites occupied by a species, is a state variable of interest in ecology. One challenge in its estimation is that detection is often imperfect in wildlife surveys. As a consequence, occupancy models that explicitly describe the observation process are becoming widely used in the discipline. These models require data that are informative about species detectability. Such information is often obtained by conducting repeat surveys to sampling sites. One strategy is to survey each site a predefined number of times, regardless of whether the species is detected. Alternatively, one can stop surveying a site once the species is detected and reallocate the effort saved to surveying new sites. In this paper we evaluate the merits of these two general design strategies under a range of realistic conditions. We conclude that continuing surveys after detection is beneficial unless the cumulative probability of detection at occupied sites is close to one, and that the benefits are greater when the sample size is small. Since detectability and sample size tend to be small in ecological applications, our recommendation is to follow a strategy where at least some of the sites continue to be sampled after first detection.

Keywords Detectability · Imperfect detection · Occupancy · Removal design · Survey design · Zero-inflated binomial

Electronic supplementary material The online version of this article (doi:10.1007/s10182-017-0292-5) contains supplementary material, which is available to authorized users.

✉ Gurutzeta Guillera-Arroita
gurutzeta.guillera@unimelb.edu.au

1 School of Biosciences, University of Melbourne, Parkville, VIC 3010, Australia

1 Introduction

Statistics play an important role in the study of ecological systems. Statistical methods are continuously refined to address important questions in ecology and are critical to support conservation and management decisions. These include techniques for estimating indices of biodiversity (Buckland et al. 2012), modelling animal movement (Patterson et al. 2008), studying population demography (open population capture–recapture; Lebreton et al. 1992), as well as estimating population abundance (e.g. closed population capture–recapture, distance sampling; Borchers et al. 2003; Buckland et al. 2001) and species occupancy (MacKenzie et al. 2006), including integrated population models (Besbeas et al. 2002). A key issue in statistical ecology methods is that ecological processes are often imperfectly observed. Hence, a common feature across methods is that they normally involve a hierarchical structure around a system process that describes the underlying biological system and an observation process that describes the data collection process (King 2014).

In this paper, we are concerned about the estimation of species occupancy, that is, the proportion of sites occupied by a species. Species occupancy is a state variable widely used in ecology (MacKenzie et al. 2006). It is often used as a metric to assess the status of populations, especially when working at large spatial scales (e.g. Wibisono et al. 2011). Occupancy plays also a fundamental role in species distribution modelling, when described as a function of environmental predictors (Franklin 2013). Detection is often imperfect in occupancy surveys, with species missed at sites they occupy (Kéry and Schmidt 2008; Yoccoz et al. 2001). Hence, when estimating species occupancy, it is important to account for species detectability (Guillera-Arroita et al. 2014b; MacKenzie et al. 2002; Lahoz-Monfort et al. 2014).

While the issue of imperfect detection has long been addressed in similar estimation problems (e.g. population abundance, Borchers et al. 2003; Buckland et al. 2001), only over the last decade has a modelling framework become widely applied to accommodate imperfect detection for species occupancy studies (MacKenzie et al. 2002; Tyre et al. 2003). Since then, there has been a rapid uptake of this general approach and many model extensions have been developed (Bailey et al. 2014). The key to the method is to collect information about the detection/non-detection of a species at sites in a way that is also informative about its detectability. Often this is achieved by carrying out several visits to each site, keeping the records from each visit separately; this is the scenario that we analyse in this paper. Alternatives include obtaining several records in one single visit through multiple simultaneous observers or spatial subsampling of sites (Guillera-Arroita 2011; Kendall and White 2009), or recording times to detection (Garrard et al. 2008; Guillera-Arroita et al. 2011). Species detection data can come from a variety of survey methods, from direct detection through visual observations (e.g. Guillera-Arroita et al. 2010a) or aural records (e.g. MacKenzie et al. 2002), to indirect detection by the location of tracks or other signs of species presence (e.g. Wibisono et al. 2011). Other innovative survey methods, such as interviews, have also been applied in this context (e.g. Zeller et al. 2011) and, more recently, surveys based on the analysis of 'environmental DNA' samples are starting to be tested and deployed, in particular for surveying aquatic species (e.g. Schmidt et al. 2013).

When confronted with designing an occupancy-detection survey, an important question is how to best allocate a given survey effort budget between the number of sites sampled and the amount of effort applied per site, to maximise the precision of the estimation of occupancy. There has been substantial work in the literature addressing this trade-off, both theoretically and via simulations (Bailey et al. 2007; Guillera-Arroita and Lahoz-Monfort 2012; Guillera-Arroita et al. 2010b, 2014c; MacKenzie and Royle 2005). Another related design question that ecologists and environmental practitioners often ask is whether occupancy-detection surveys need to continue after first detection at a site. Further records after the first detection do not add directly any information about the occupancy status of sites, so they could seem unnecessary. However, those additional records can help in the estimation of species detectability and, through this, lead to a better estimation of species occupancy across the landscape. The question is thus whether it is worth spending survey effort at a site after the species is first detected there, or whether one should dedicate that effort to surveying new sites (i.e. is a so-called 'removal design' superior to a 'standard design'?; Fig. 1). To our knowledge, this question has only been very briefly addressed in the literature. MacKenzie and Royle (2005) used asymptotic results to assess the relative efficiency of an *optimal* removal design with respect to an *optimal* standard design under *constant* occupancy and detectability. Based on this analysis, they concluded that there was "strong evidence that a removal design is much more efficient than a standard design for estimating occupancy" if detection probability is constant. However, they noted that they expected the data yielded by removal designs to provide less flexibility for modelling, particularly for exploring potential sources of variation in detection probability, and hence suggested that a removal design could be less robust in general.

In this paper, we draw on asymptotic theory and Monte Carlo simulations to evaluate the relative merits of the standard and removal design strategies in occupancy-detection models across a range of scenarios. All our evaluations are presented from the point of view of maximum-likelihood estimation. We organize the paper as follows: in Sect. 2, we provide an overview of the basic occupancy-detection model and some of its key properties; in Sect. 3 we compare the standard and removal design strategies under a scenario of constant occupancy and detectability; in Sect. 4, we extend the comparison to spatial heterogeneity in detection probability and in Sect. 5 we explore scenarios where such heterogeneity is explicitly accounted for in the model. We conclude with a summary of our key findings. Ecologists interested in the topic, but less inclined to follow the statistical details of our evaluation, can refer directly to that concluding section.

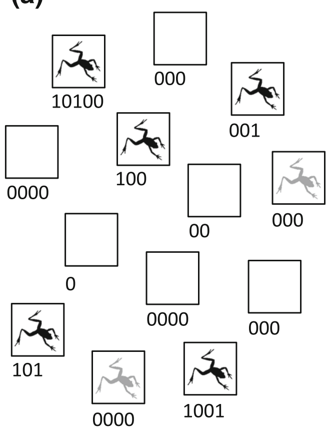
2 Occupancy-detection model

In its basic form, the occupancy-detection modelling approach assumes independence in the occupancy status of sites and describes the detection/non-detection records of the species at occupied sites as the outcome of a series of independent Bernoulli trials (MacKenzie et al. 2002; Tyre et al. 2003). Other key assumptions are the absence of false positive records and the closure of sites to changes in occupancy status (i.e. each site is either occupied or unoccupied, and remains so for the whole survey period). The likelihood function corresponding to this model can be written as

$$L(\boldsymbol{\psi}, \boldsymbol{p}) = \prod_{i=1}^{S} \left\{ \psi_i p_i^{d_i} (1 - p_i)^{K_i - d_i} + I(d_i = 0)(1 - \psi_i) \right\}, \tag{1}$$

where ψ_i is the probability that the species occurs at site i, p_i is the probability that the species is detected during a survey visit to site i *conditional* on it being occupied (hereafter, simply called detectability), S is the number of sites sampled, K_i is the number of survey visits carried out at site i, and d_i is the number of detections at the site. This is a zero-inflated binomial model, where sites with no detections ($d_i = 0$) can either arise because the species is not present or because the species was present, but missed in all survey visits. Normally detections are coded as '1' and non-detections as '0' (Fig. 1), and this is what we assume hereafter.

(a)



(b) Standard design

site	visit 1	2	3	…	K
1	**1**	**1**	0	…	0
2	**1**	0	0	…	0
3	0	0	0	…	0
4	**1**	0	**1**	…	0
5	0	0	0	…	0
6	0	0	0	…	0
7	**1**			…	
8	0	0	0	…	**1**
…	…	…	…	…	…
S	0	0	**1**	…	0

Sites are surveyed a predefined number of times

Removal design

site	visit 1	2	3	…	K
1	**1**	--	--	…	--
2	**1**	--	--	…	--
3	0	0	0	…	0
4	**1**	--	--	…	--
5	0	0	0	…	0
6	0	0	0	…	0
7	**1**	--	--	…	--
8	0	0	0	…	**1**
9	…	…	…	…	…
10	0	0	**1**	…	--
11	0	0	0	…	**1**
12	0	0	0	…	0
…	…	…	…	…	…
S'	0	**1**	--	…	--

More sites can be visited, for a survey budget

Surveys stop if the species is detected

Fig. 1 Occupancy-detection survey data. The diagram in **a** exemplifies a typical occupancy-detection study, where a number of sites are surveyed with replication. Sampling sites can be naturally discrete habitat features (e.g. ponds) or may be defined as artificial units over a continuous habitat (e.g. forest quadrats). False negative records are likely, given that species detection is imperfect. Hence detection histories with all zeros can come from sites with or without the species. The tables in **b** exemplify occupancy-detection data sets collected under two design strategies: a 'standard design', where all sites are surveyed a predefined number of times; and a 'removal design', where sites are surveyed until the species is detected, and/or until there have been a maximum number of visits. The term 'removal design' is used to reflect that sites are removed from the pool of sites being sampled once the species is detected, and by analogy with removal studies on animal populations (MacKenzie and Royle 2005)

We consider now the two design strategies under evaluation: (1) the 'standard design', where all sites are surveyed K_M times, and (2) the 'removal design', where sites are surveyed up to a maximum of K_M times or until the species is detected, whatever comes first (Fig. 1b). Here we assume that K_M is equal for all sites, but this is not necessary in practise. Note that the number of actual visits carried out at a site $i(K_i)$ is a random variable under the removal design. It follows a distribution that is a mixture of K_M (the number of visits carried out at an empty site) and a modified geometric distribution (that describes the number of visits required to get a detection at occupied sites, limited to K_M). The expected number of visits at a site i under a removal design is thus

$$\begin{aligned}\mathbb{E}\left[K_i\right] &= (1-\psi_i)\, K_M + \psi_i \left\{\sum_{k=1}^{K_M-1} k p_i\,(1-p_i)^{k-1} + K_M\,(1-p_i)^{K_M-1}\right\}\\ &= (1-\psi_i)\, K_M + \psi_i\, p_i^*/p_i,\end{aligned} \tag{2}$$

where we use p_i^* to denote the cumulative conditional probability of detection at a site after K_M visits, i.e. $p_i^* = 1-(1-p_i)^{K_M}$. Note that, as the amount of replication increases and the chances of detecting the species at an occupied sites approach one ($p_i^* \approx 1$), the expected number of visits at an occupied site approaches $1/p_i$ and the distribution becomes geometric.

Let us assume that occupancy and detection are constant to obtain some simple, but useful results. The likelihood function in (1) can now be written as

$$L\,(\psi, p) = \psi^{S_d} p^{n_1}\,(1-p)^{n_0-(S-S_d)K_M}\left(1-\psi p^*\right)^{S-S_d}, \tag{3}$$

where S_d is the number of sites where the species is detected, n_1 is the total number of detections (i.e. the number of ones in the dataset), and n_0 is the total number of surveys that lead to no detection (i.e. the number of zeros in the dataset). The first three terms of the product in (3) represent the contribution of the S_d sites where the species has been detected at least once. The fourth term corresponds to the contribution of the $S - S_d$ sites without detection, for which there is ambiguity about their true occupancy status (i.e. the species might have been absent or, alternatively, present, but missed in all survey visits). We can write $n_0 = K_T - n_1$, where $K_T = \sum_{i=1}^{S} K_i$ is the total number of surveys carried out. In the standard design, we know a priori that $K_T = K_M S$, hence (S_d, n_1) are sufficient statistics in (3). In the removal design, there is by design at most one detection per site so $n_1 = S_d$ and hence (S_d, K_T) are sufficient statistics. By doing a transformation $\theta = \psi p^*$ (Morgan et al. 2007), we can simplify the likelihood in (3) into two independent components as follows:

$$L\,(\psi, p) = \left\{\theta^{S_d}\,(1-\theta)^{S-S_d}\right\}\left\{\left(\frac{1-p^*}{p^*}\right)^{S_d}\left(\frac{p}{1-p}\right)^{n_1}(1-p)^{K_T-K_M S}\right\}.$$

Differentiating and solving the two resulting equations gives the maximum-likelihood estimators (MLEs)

$$\hat{\psi} = \frac{S_d}{S\hat{p}^*}, \quad \text{and} \quad \frac{n_1}{\hat{p}} - \frac{K_M S_d}{\hat{p}^*} = K_T - K_M S. \tag{4}$$

Since the right hand side of the second expression is zero under a standard design, then $\frac{\hat{p}}{\hat{p}^*} = \frac{n_1}{K_M S_d}$. The expressions in (4) are only valid under certain conditions. Generalizing conditions given by Guillera-Arroita et al. (2010b) for the standard design, we find that (4) applies when

$$\left(\frac{S - S_d}{S}\right) \geq \left(\frac{n_0}{K_T}\right)^{K_M},$$

that is, when the proportion of sites without detection is greater than the proportion of zeros in the dataset to the power of K_M (see the supplementary material for the derivation). Otherwise, the MLEs lie at the boundary $\hat{\psi} = 1$ with $\hat{p} = n_1/n_0 + n_1$. This suggests that the estimation can be poor when proportionally there are few ones in total in the sites where the species is detected, especially when K_M is small.

We can now compute the asymptotic variance-covariance matrix under both designs by inverting the expected Fisher information matrix, which has elements

$$j_{ik} = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_k}\right],$$

where $\mathcal{L} = \log(L)$ is the log-likelihood function with L given by (3). Assuming no model violations, we have that $\mathbb{E}[S_d] = S\psi p^*$, $\mathbb{E}[n_1] = S\psi K_M p$ under a standard design, and $\mathbb{E}[K_T] = S\mathbb{E}[K_i] = S\{(1-\psi)K_M + \psi p^*/p\}$ under a removal design. The asymptotic variances of the occupancy and detectability estimators under a standard design are (Guillera-Arroita et al. 2010b)

$$\begin{aligned}
\text{var}\left(\hat{\psi}\right) &= \frac{\psi}{S}\left\{(1-\psi) + \frac{(1-p^*)}{p^* - K_M p(1-p)^{K_M-1}}\right\}, \\
\text{var}\left(\hat{p}\right) &= \frac{p(1-p)}{SK_M\psi}\left\{\frac{p^*}{p^* - K_M p(1-p)^{K_M-1}}\right\}, \\
\text{covar}\left(\hat{\psi}, \hat{p}\right) &= \frac{-p}{S}\left\{\frac{(1-p^*)}{p^* - K_M p(1-p)^{K_M-1}}\right\},
\end{aligned} \tag{5}$$

and under a removal design

$$\begin{aligned}
\text{var}\left(\hat{\psi}\right) &= \frac{\psi}{S}\left\{(1-\psi) + \frac{(1-p^*)p^*}{p^{*2} - K_M^2 p^2 (1-p)^{K_M-1}}\right\}, \\
\text{var}\left(\hat{p}\right) &= \frac{p(1-p)}{SK_M\psi}\left\{\frac{p^* K_M p}{p^{*2} - K_M^2 p^2 (1-p)^{K_M-1}}\right\},
\end{aligned} \tag{6}$$

$$\operatorname{covar}\left(\hat{\psi},\hat{p}\right)=\frac{-p}{S}\left\{\frac{(1-p^{*})\,K_M p}{p^{*2}-K_M^2 p^2 (1-p)^{K_M-1}}\right\}.$$

We use the expressions of the asymptotic variance of the occupancy estimator in the next section to compare the efficiency of the two design strategies. As we are interested in making comparisons where the same total survey effort (E) is applied to both designs, we can re-express (5) and (6) as a function of E. Let us assume that the cost of survey visits is equal across sites, and that the first visit to each site is C_1 times more costly than successive visits (e.g. due to access costs). Under a standard design, disregarding the rounding, the number of sites surveyed given a total effort E is $S = E/(C_1 + K_M - 1)$. Under a removal design, the *expected* number of sites surveyed is $\mathbb{E}[S] = E/(C_1 + \mathbb{E}[K_i] - 1)$, where the expected number of visits to a site is $\mathbb{E}[K_i] = (1-\psi)\,K_M + \psi p^*/p$. By replacing S in (5) and $\mathbb{E}[S]$ in (6) accordingly, we arrive at expressions for the variance of the occupancy estimator that are only a function of ψ, p, K_M and E. From these expressions, we can also identify the optimal amount of replication (optimal K_M) to minimize the variance of the occupancy estimator under each design strategy (MacKenzie and Royle 2005); since E is just a scaling factor in the variance expressions, the optimal K_M is only a function of ψ and p (always assuming large sample size). We find that the maximum survey effort required under an *optimal* removal design is always greater than under an optimal standard design (Table 1). However, this effort is only to be fully applied to sites where the species is not detected; in practise, the average effort applied per site ($\mathbb{E}[K_i]$) is often smaller compared to that in an optimal standard design, unless both occupancy and detection probability are small.

3 Performance when occupancy and detectability are constant

In this section we evaluate the relative efficiency of the removal and standard designs assuming constant occupancy and detectability, across different combinations of values of these two parameters (from 0.1 to 0.9). Here and elsewhere we focus on the precision of the occupancy estimator ($\hat{\psi}$). In all comparisons, the same total effort is applied under both designs. We compute relative efficiency as the ratio between the root mean square error (RMSE) of $\hat{\psi}$ under the standard design and the RMSE under the removal design: $\text{eff} = \text{RMSE}\left(\hat{\psi}_S\right)/\text{RMSE}\left(\hat{\psi}_R\right)$. We consider the case where the optimum level of replication is applied (as per Table 1), and scenarios where a given amount of replication is applied regardless of its optimality. We first do an assessment assuming large samples by considering the asymptotic variance expressions in (5) and (6). We then complement this assessment by running simulations to check whether the patterns observed change as the sample size (E) is reduced. We simulate 10,000 datasets for each occupancy-detection scenario considering that the survey effort is limited to $E = 500$ cost units. We also run simulations for a set of selected scenarios, reducing survey effort from 2000 units down to 200 units by decrements of 200. All our simulations are run in R version 3.1.1 (R Development Core Team 2016), with parameter estimates obtained via maximum-likelihood estimation. We use the

Table 1 Optimum amount of replication to maximize the precision of the occupancy estimator for a given budget under the standard and removal designs as a function of occupancy and detectability

(a)

	Opt. K_M (standard)									Opt. K_M (removal)									Opt. $\mathbb{E}[K_i]$ (removal)								
p \ ψ	.1	.2	.3	.4	.5	.6	.7	.8	.9	.1	.2	.3	.4	.5	.6	.7	.8	.9	.1	.2	.3	.4	.5	.6	.7	.8	.9
.1	14	15	16	17	18	20	23	26	34	23	24	25	26	28	31	34	39	49	22	21	20	19	19	18	17	16	14
.2	7	7	8	8	9	10	11	13	16	11	11	12	13	13	15	16	19	23	10	10	10	10	9	9	8	8	7
.3	5	5	5	5	6	6	7	8	10	7	7	7	8	8	9	10	12	15	7	6	6	6	6	6	5	5	4
.4	3	4	4	4	4	5	5	6	7	5	5	5	6	6	6	7	8	10	5	4	4	5	4	4	4	4	3
.5	3	3	3	3	3	3	4	4	5	4	4	4	4	4	5	5	6	8	4	4	3	3	3	3	3	3	3
.6	2	2	2	2	3	3	3	3	4	3	3	3	3	3	4	4	5	6	3	3	3	2	2	3	2	2	2
.7	2	2	2	2	2	2	2	3	3	2	2	2	3	3	3	3	4	5	2	2	2	2	2	2	2	2	2
.8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	4	2	2	2	2	2	2	2	2	2
.9	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	1	1	1	1

(b)

	Opt. K_M (standard)									Opt. K_M (removal)									Opt. $\mathbb{E}[K_i]$ (removal)								
p \ ψ	.1	.2	.3	.4	.5	.6	.7	.8	.9	.1	.2	.3	.4	.5	.6	.7	.8	.9	.1	.2	.3	.4	.5	.6	.7	.8	.9
.1	16	17	18	19	20	22	25	28	35	24	25	26	28	30	32	36	41	51	23	22	21	21	20	19	18	16	14
.2	9	9	9	10	11	11	12	14	17	12	12	13	14	15	16	18	20	25	11	11	11	10	10	9	9	8	7
.3	6	6	6	7	7	8	8	9	11	8	8	9	9	10	10	12	13	16	8	7	7	7	7	6	6	5	5
.4	5	5	5	5	5	6	6	7	8	6	6	6	7	7	8	8	10	12	6	5	5	5	5	5	4	4	3
.5	4	4	4	4	4	4	5	5	6	4	5	5	5	5	6	6	7	9	4	4	4	4	3	4	3	3	3
.6	3	3	3	3	3	4	4	4	5	3	4	4	4	4	5	5	6	7	3	4	3	3	3	3	3	3	2
.7	2	3	3	3	3	3	3	3	4	3	3	3	3	3	4	4	5	6	3	3	3	2	2	2	2	2	2
.8	2	2	2	2	2	2	3	3	3	2	2	2	3	3	3	3	4	4	2	2	2	2	2	2	2	2	2
.9	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	2	2	2	2	2	1	1	1	1

In (a) the cost of all surveys is equal ($C_1 = 1$); in (b) the first visit to a site is five times more costly than successive visits ($C_1 = 5$). Under a standard design, all sites are surveyed K_M times. Under a removal design, K_M is the maximum number of surveys to carry out at a site without detection. The expected number of surveys carried out under the optimal removal design is shown rounded in the rightmost panel of the tables. In this panel, the line delimits scenarios where the optimal removal design, in average, requires per site one visit more than the optimal standard design. These results are based on asymptotic approximations and also partly available in MacKenzie and Royle (2005)

function *optim* to find the maximum of the likelihood function with the Nelder–Mead simplex search algorithm (Nelder and Mead 1965).

As in MacKenzie and Royle (2005), our asymptotic results indicate that the removal design is more efficient than the standard design in most scenarios under *optimal* conditions (Fig. 2a, left). The removal design is worse when occupancy and detection probabilities are small. In those cases, performance differences are relatively moderate. However, when we consider other amounts of replication, we observe that the removal design is substantially worse than the standard design when K and p are low (Fig. 2a, centre). This indicates that the removal design is more sensitive than the standard design to reductions in the cumulative detectability level (p^*), as illustrated also by Fig. 3a. Conversely, when K and p are large, the removal design outperforms the standard design (Fig. 2a, right). This is because, when cumulative detectability is close to perfect ($p^* \approx 1$), modelling the detection process is not relevant for the estimation of ψ and, by stopping surveys after the first detection, the removal design avoids "wasting" effort in revisiting sites where the species is already detected. If

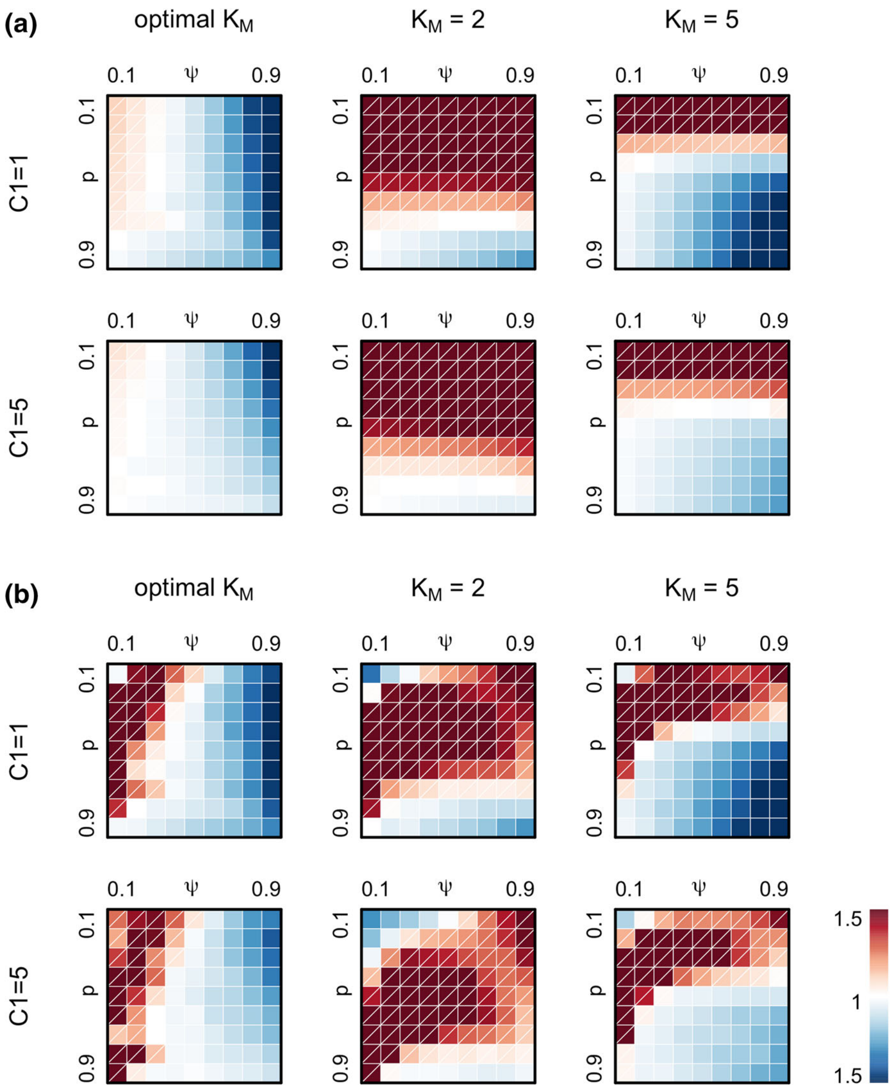


Fig. 2 Relative efficiency of standard vs. removal design under a range of scenarios of constant occupancy (ψ) and detection (p) probabilities. *Red cells* (with a diagonal line) indicate conditions when the removal design is less precise. *Each subplot* corresponds to one of three scenarios of replication (optimal for each case and type of design, fixed to $K_M = 2$, or fixed to $K_M = 5$), and to one of two scenarios of cost of first survey ($C_1 = 1$ or 5). In **a** results are derived from asymptotic approximations; in **b** results are obtained via simulation (10,000 runs) for a total effort E of 500 units ('optimal K_M' still determined based on large sample approximations). The quantity represented is the ratio of the RMSEs of the occupancy estimator. For representation purposes, the largest RMSE is always divided by the smallest and the colour range covers ± 1.5 (values outside this range are clipped to ± 1.5)

detection is close to perfect ($p^* \approx 1$), the precision of the occupancy estimator (now simply the estimator of a binomial proportion) is proportional to the number of sites sampled, and this number is greater under the removal design (Fig. 3b). The benefits

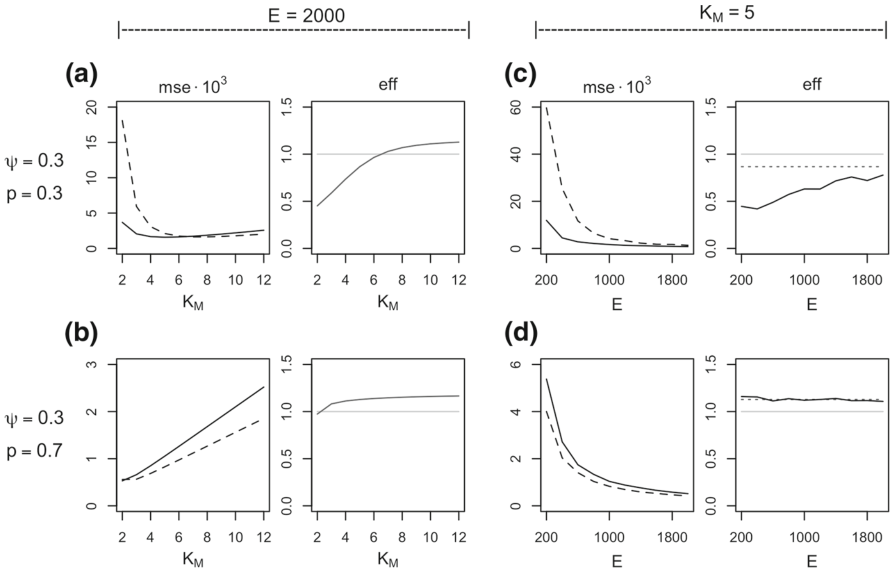


Fig. 3 Comparison of the performance of the occupancy estimator obtained from a standard design and a removal design, for two selected scenarios of occupancy (ψ) and detectability (p). The *left plot* in *each pair* represents the performance of each design strategy separately in terms of MSE: *solid line* standard design, *dashed line* removal design. The *right plot* in *each pair* shows relative efficiency, the ratio RMSE standard/RMSE removal (hence values <1 indicate that the standard design is best). The *left pairs* (**a**, **b**), obtained based on asymptotic approximations, show how the relative efficiency depends on the amount of replication K, with the removal design being particularly sensitive to working with less replication than optimal. The *right pairs* (**c**, **d**), obtained from simulations (10,000), illustrate how the removal design can be more sensitive to reductions in sampling effort (E). Here the efficiency plots include a *dotted line* showing the efficiency as calculated based on asymptotic approximations. In all cases it is assumed that the cost of the first survey $C1$ is 1. A more extensive set of results is presented as supplementary material

of the removal design are reduced as the cost C_1 of sampling a new site increases (there is less blue in the second row in Fig. 2a). This is expected because, the higher the C_1, the fewer survey visits can be made with the survey effort that is redirected to a new site.

Our simulations reveal that the removal design is more severely affected by reductions in survey effort (E). As the sample size decreases and we move away from conditions where asymptotic approximations hold, the number of scenarios of (ψ, p) where the removal design underperforms the standard design increases (Fig. 2b). In principle we would expect both designs to degrade at a similar rate, with the variance increasing proportionally as E decreases (Fig. 3d), and hence the relative efficiency to remain constant. However, as large sample approximations start to break down, we find that the performance of the removal design tends to degrade faster with reductions of sampling effort (Fig. 3c). The situation changes when both p^* and E are sufficiently small. Then we find that the MSE of the standard design appears to increase at a higher rate with reductions in E (see results of detailed simulations in the supplementary material). However, the MSE of the standard design only becomes higher than that of the removal design in situations where the performance of both estimators is so poor that there is little practical value in the estimation. These cases correspond to the blue cells in the upper left corner of Fig. 2b.

4 Performance when detectability is heterogeneous

In this section we compare the performance of the two designs under scenarios where there is spatial heterogeneity in detectability. Detectability can vary between sites for a number of reasons, including differences in abundance or differences in visibility due to specific habitat characteristics. Detectability many also vary between sites if there are several observers conducting surveys, but there is no randomization, with each observer carrying out all visits to a site. Here we consider the case in which, although detectability varies across sites, this is not accounted for in the modelling. In our evaluation, we assume that site-specific detectability follows a beta distribution, $p_i \sim \text{beta}\,(c, d)$ where c and d are the shape parameters of the distribution and are strictly positive real numbers. It is known that site heterogeneity in p can induce bias in the estimation of ψ (Royle 2006). Assuming large sample size, we can study the bias introduced by solving the expressions in (4) for the expected values of the data under heterogeneity (either (S_d, n_1) or (S_d, K_T), depending on the design). We obtain the expectations we need ($\mathbb{E}\,[p]\,, \mathbb{E}\left[p^*\right], \mathbb{E}\left[p^*/p\right]$) considering that for the beta distribution above $\mathbb{E}\,[p] = \frac{c}{c+d}$ and $\mathbb{E}\left[p^x\right] \approx \frac{c+x-1}{c+d+x-1}\mathbb{E}\left[p^{x-1}\right]$ (Welsh et al. 2013). Guillera-Arroita et al. (2014b) present asymptotic results for the bias induced under a standard design across different levels of heterogeneity. Practically the same results are obtained when we consider a removal design (Fig. 4), with the bias being in some cases slightly smaller than that in the standard design. As could be expected, heterogeneity in detectability does not introduce bias in the occupancy estimator when cumulative detectability p^* is close to one (large mean p or K_M), as then the estimation of the detection probability has little influence in the estimation of occupancy probability. When detection is imperfect ($p^* < 1$), the greater the heterogeneity in detectability, the greater is the bias induced in the occupancy estimator. As the amount of bias decreases with K, we can expect the optimal levels of replication to be larger than those derived in Table 1.

We simulate some selected scenarios to explore further the performance of both designs under heterogeneity in detectability. We test all combinations of ψ and mean p taking values 0.3, 0.5 and 0.7, for two levels of replication ($K_M = 2$ and 5), three levels of survey effort ($E = 5000$, 1000 and 500), and encompassing the amounts of heterogeneity considered in Fig. 4. The simulation results show no general advantage of one design over the other based on their ability to deal with unmodelled heterogeneity in detectability (Fig. 5; full simulation results are provided in the supplementary material). Heterogeneity does not have an effect when cumulative detectability (p^*) is close to perfect (Fig. 5a), which is when a removal design is more efficient than a standard design. Otherwise, heterogeneity degrades the performance of the occupancy estimator under both designs and does so in a similar way (but see an exception below). Where there is substantial heterogeneity and the sample size is very large, the bias induced by heterogeneity dominates the performance of the occupancy estimator. Then, the relative efficiency of the designs approaches one (Fig. 5b), or is slightly bet-

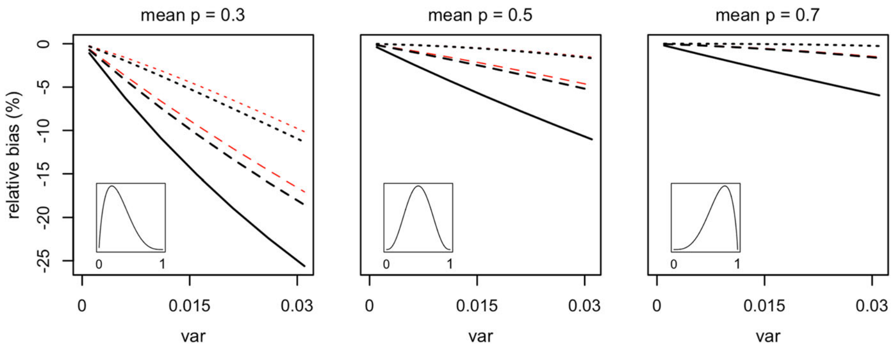


Fig. 4 Asymptotic bias in the occupancy estimator for the standard and removal design for different levels of unmodelled heterogeneity in detection probability p. The x axis represents the variance of the beta distribution used to characterize detectability. *Curves* are shown for three levels of replication: $K_M = 2$ (*solid line*), $K_M = 5$ (*dashed*), $K_M = 10$ (*dotted*). *Red lines* (thinner) correspond to the removal design and *black lines* (thicker) to the standard design. Results coincide for both designs when $K_M = 2$. *Insets* show the density function of the beta distribution that describes detectability for the case of the maximum variance contemplated (0.03)

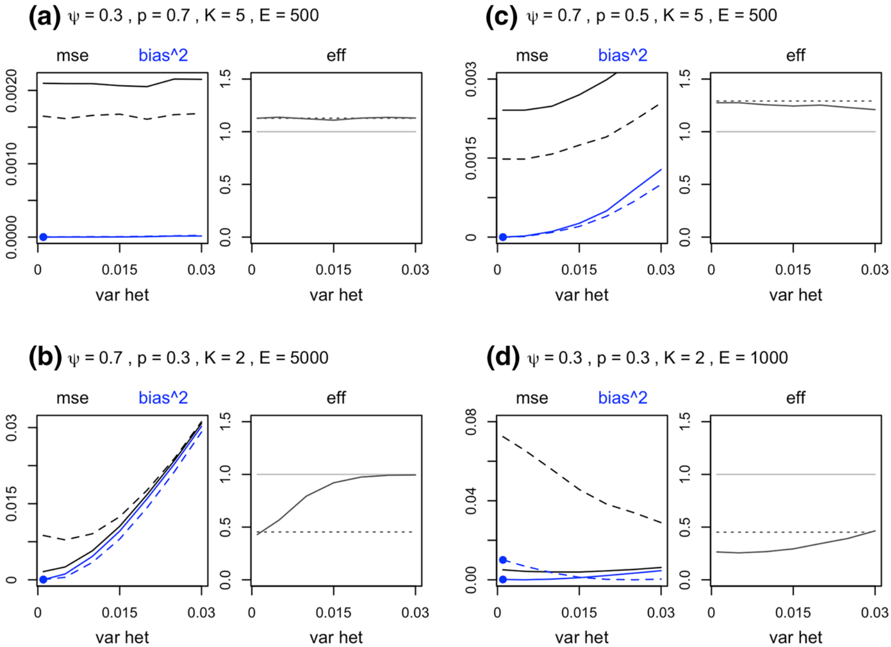


Fig. 5 Comparison of the performance of the occupancy estimators obtained from a standard design and a removal design under increasing spatial heterogeneity in detectability, for four selected scenarios. The *left plot* in *each pair* represents the performance of each design strategy separately: *black lines* MSE, *blue lines* bias squared (marked also with a dot); *solid lines* standard design, *dashed lines* removal design. The *right plot* in *each pair* shows relative efficiency, the ratio RMSE standard/RMSE removal (hence values <1 indicate that the standard design is best, as in Fig. 3). In the efficiency plots, the *dotted line* shows for reference the efficiency as calculated based on asymptotic approximations and assuming no heterogeneity. Results obtained from simulations (5000 per case)

ter for the removal design as then the induced bias can be somewhat smaller (Fig. 4). For smaller sample sizes, the variance also plays an important role in estimation performance. In these cases, the *relative* efficiency of the designs may remain practically constant regardless of the heterogeneity (Fig. 5c), or it may shift towards the removal design, as its performance appears to improve with respect to the case without heterogeneity (Fig. 5d). Such improvements, which at first glance seem counterintuitive, are explained because the shift towards smaller estimated values induced by heterogeneity can lead to a decrease in variance (and bias) in cases where ψ, p and K are small (see an example in the supplementary material). These are cases where the removal design is worse than the standard design and, despite these improvements, it remains so across the heterogeneity levels tested. In summary, our simulation results indicate that, when spatial heterogeneity in detectability is unaccounted for, the standard design continues outperforming the removal design when cumulative detectability is not close to perfect, but that the gap in performance decreases as the amount and impact of heterogeneity increases.

5 Ability to model heterogeneity in detectability

In this section we consider again that detectability varies spatially, but we now assume that predictors are available to describe this heterogeneity and that these predictors are incorporated in the modelling. Our aim with this assessment is to explore whether we find evidence that, as hinted elsewhere (MacKenzie and Royle 2005), a removal design could tend to perform worse that a standard design in this setting due to a more limited ability to characterize the detection process well. To this purpose, we run simulations where detectability at each site i is generated as a linear combination of a number N of covariates through a logit link function: $\text{logit}\,(p_i) = \beta_0 + \sum_{n=1}^{N} \beta_n C_{n_i}$. Here, spatial heterogeneity results from the combined effect of the spatially varying covariates (C_n). We generate values for these N predictors by drawing from independent standard normal distributions and we keep all regression coefficient equal, i.e. $\beta_n = \beta, n = 1 \ldots N$. We compare the performance of the two designs for a given amount of heterogeneity (detailed next) across variable degrees of complexity in the detection process (here captured by the number of predictors). We achieve this by setting $\beta = \sqrt{h/N}$, where h is the variance of the linear predictor ($\beta \sum_n C_{n_i}$) and hence represents the amount of heterogeneity in p; this follows because in our simulations $\text{logit}\,(p_i) \sim N\left(\beta_0, \beta^2 N\right)$, so $h = \beta^2 N$. We set $h = 1.28$, which corresponds to a variance in detectability of about 0.05, and assess performance when $N = 2$, 5, 10 or 15 predictors. To avoid convergence problems when fitting the more complex models, we increase the maximum number of iterations allowed by the R optimization function *optim*. We run each assessment for a range of scenarios of occupancy ($\psi = 0.3, 0.5, 0.7$), detectability in a visit ($\tilde{p} = 0.3, 0.5, 0.7$, where $\tilde{p}$ represents detection probability at "typical" sites, that is, those where all covariates take value zero, i.e. $\tilde{p} = \left(1 + \mathrm{e}^{-\beta_0}\right)^{-1}$) and replication ($K = 2, 3, 5$ visits). We run 1000 simulations for each combination of parameters, and compute the RMSE of the occupancy estimator under both designs. In all cases we assume $E = 500$. For reference, we include in the assessment the case where heterogeneity is unaccounted for in the model.

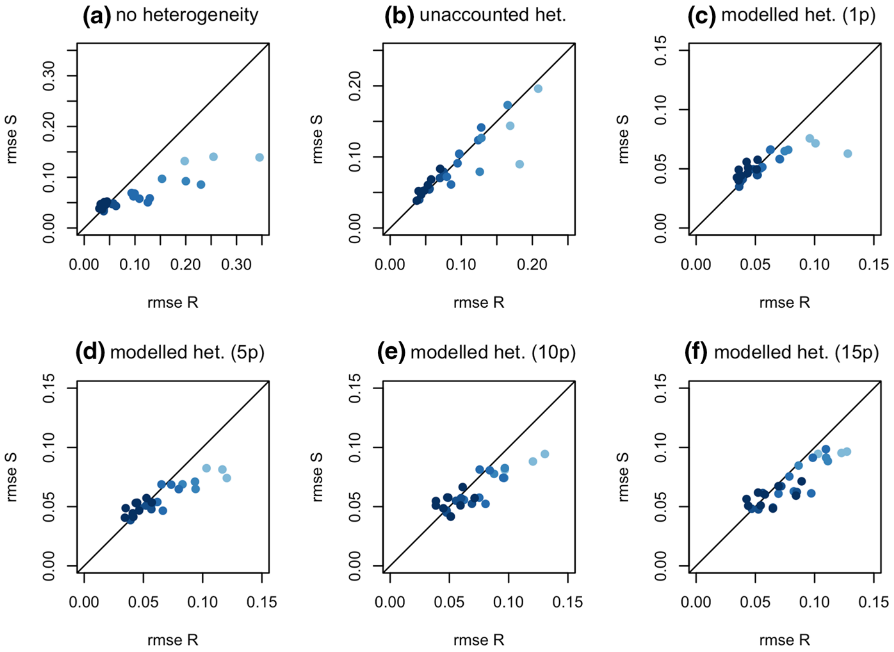


Fig. 6 Comparison of the performance of the occupancy estimators obtained from a standard design (S) and a removal design (R) under increasing complexity in the detection process (number of predictors describing the detectability, in *brackets*). *Each dot* within a *subplot* represents a scenario of occupancy, detectability and amount of replication (as detailed in the main text). Except for **a**, in all scenarios the same amount of spatial heterogeneity in detectability is applied (variance $= 1.28$ on the logit scale). Survey effort is $E = 500$ in all cases. In **b** heterogeneity is not modelled. In **c**–**f** heterogeneity is modelled as a function of the full set of predictors that describe it (which varies from 1 to 15, depending on the subplot). Performance is measured as root mean square error (RMSE). *Colour range* represents general detectability level in each scenario, quantified as the cumulative probability of detection at a "typical" site where all covariates take value zero, i.e. $\tilde{p}^* = 1 - (1 - \tilde{p})^{K_M}$. Note the different scales in the axes among subplots

The results of these simulations (Fig. 6) demonstrate again that both designs lead to similar performance when there is substantial heterogeneity in detection that is not modelled. We see this in the scatterplot in Fig. 6b, where the points are largely along the diagonal (each point in the plot represents the RMSEs of a standard and a removal design for a given scenario). The removal design tends to be best when the cumulative probably of detection p^* is high and having information about the detection process thus plays a minor role (darker points on the lower part of the plot). On the other hand, the removal design is more likely to underperform when the cumulative detectability p^* is low (lighter points of the right side of the plot). When heterogeneity is due to a single covariate, and hence relatively easy to model, we find that both methods improve their performance when the covariate is incorporated, although the removal design is still prone to problems when detectability is low (lighter points in Fig. 6c). As the detection process becomes more complex (i.e. heterogeneity is generated by the combined effects of many covariates), it becomes more difficult to model, which implies poorer performance for the occupancy estimator (Fig. 6d–f). Our results seem to suggest that the removal design underperforms in these situations, as the number of points (scenarios) under the diagonal becomes greater. This is in line with the original expectation.

6 Discussion

We have assessed and compared the performance of two general survey design strategies for estimating species occupancy under imperfect detection. We set out to answer the practical question of whether surveys should continue at a site after first detection, or whether instead it is more valuable to redirect the remaining survey effort to sampling more sites. Answering such a question in full detail is difficult as there are many scenarios one could evaluate, and the number of combinations grows as more realism (and hence complexity) is added to the evaluation. Our work is certainly not a full account of all possible situations, but we aimed to cover some important aspects of this type of study to identify general patterns about how each of the two survey design strategies performs comparatively. We have not considered cases where occupancy is modelled as a function of environmental variables, but we expect that the behaviours observed for the occupancy estimator are also likely to apply qualitatively when making inference about environmental relationships.

Our investigations deal with a setting where the study's objective is estimating occupancy probabilities, which relate to the *proportion* of sites occupied by the species across an area. This is not to be confused with cases where "occupancy surveys" are carried out with the aim of establishing the actual occupancy *status* of specific sites (i.e. whether the species is present or not; e.g. Kéry 2002). In such cases, where no estimation of the detection process is involved, there is in principle no point in continuing surveys after the species is first detected at a site. Where such surveys are to be applied across multiple sites, a design question is how to allocate a given survey effort budget to maximize the number of sites with detection, or to optimize the outcome with respect to a related objective function (e.g. minimize total costs in an invasive species eradication programme; Guillera-Arroita et al. 2014a).

6.1 Key findings

We compared the performance of the standard and removal design strategies for a fixed total survey effort focusing on the RMSE of the occupancy estimator. Our key findings are as follows (we use R and S to refer to the removal and standard designs, respectively):

- *Optimal conditions (constant probabilities, large sample size, optimal replication)*: R is in general more efficient than S when the optimal amount of replication per site is used, except when occupancy and detection probabilities are small.
- *Departures from optimal replication*: R is more sensitive than S to using less replication per site than the optimum level, and tends to underperform in these situations. R is more efficient when the chosen replication is such that cumulative detectability at a site (p^*) is close to 1.

- *Small sample size:* R appears less robust than S to reductions in sample size; its performance tends to degrade faster as the total sampling effort decreases.
- *Spatial heterogeneity in detectability (unmodelled)*: Both R and S tend to degrade in performance when there is heterogeneity in detectability (unless p^* is close to 1), and they do so in a similar way because the bias induced by the heterogeneity is almost the same. Where there is substantial heterogeneity that is not accounted for and the bias dominates, we can expect both designs to perform similarly. In other words, S continues outperforming R when overall detectability is not close to perfect, but the performance gap narrows down as the amount and impact of unaccounted heterogeneity increases.
- *Complex detection process (spatial heterogeneity in detectability that is modelled)*: The performance of the occupancy estimator degrades as the detection process becomes more complex and hence difficult to model. It appears that this effect might be more prominent when R is used.

6.2 Survey design recommendations

Wildlife managers and researchers planning surveys to estimate species occupancy probabilities accounting for imperfect detection are faced with the decision of whether to use a standard or a removal design strategy. From our study, we can conclude that a removal design is better when overall detectability at the sites is close to perfect. This happens when detectability at a single visit, p, and the maximum number of visits per site, K_M, are such that $p^* = 1 - (1 - p)^{K_M} \approx 1$ (but note that a design with excessive replication can be far from optimal and hence not necessarily desirable). In such cases, imperfect detection is not an issue with the survey effort applied to each site and the estimation of occupancy will see limited benefit from characterizing the detection process well; more can be gained by visiting more sites. In contrast, a standard design appears more robust to working with low detection probabilities, especially when the sample size is small and in situations where the description of the detection process is complex (e.g. many predictors involved). Since these conditions are more challenging and likely to apply in real ecological studies, our suggestion is to follow a strategy where at least some of the sites continue to be sampled after first detection. Although this implies sampling fewer sites overall, the extra information collected at sites where the species is confirmed can lead to better estimation of occupancy through a better estimation of the detection process. This recommendation is in line with suggestions by MacKenzie and Royle (2005), who propose to, at least partially, follow the standard design strategy as a likely means to gain robustness and "flexibility for modelling the collected data".

To conclude we note that our evaluation was carried out with the ecological context in mind, but our findings can be of use more generally whenever a binomial proportion needs to be estimated and the success category is imperfectly observed. We suspect this type of estimation problem to arise as well in other fields. These disciplines can gain from reference to the ecological statistics literature, which is rich on this topic and has expanded from inference about occupancy patterns, as considered in this paper, to covering the modelling of occupancy dynamics and systems with multiple occupancy states.

Acknowledgements This work was supported by the Australian Research Council (ARC) Centre of Excellence for Environmental Decisions. The authors thank Byron Morgan for comments on an earlier version of the manuscript.

References

Bailey, L.L., Hines, J.E., Nichols, J.D., MacKenzie, D.I.: Sampling design trade-offs in occupancy studies with imperfect detection: examples and software. Ecol. Appl. **17**(1), 281–290 (2007)

Bailey, L.L., MacKenzie, D.I., Nichols, J.D.: Advances and applications of occupancy models. Methods Ecol. Evol. **5**(12), 1269–1279 (2014)

Besbeas, P., Freeman, S.N., Morgan, B.J.T., Catchpole, E.A.: Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. Biometrics **58**(3), 540–547 (2002). doi:10.1111/J.0006-341x.2002.00540.X

Borchers, D.L., Buckland, S.T., Zucchini, W.: Estimating Animal Abundance: Closed Populations. Springer, New York (2003)

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L.: Introduction to Distance Sampling. Oxford University Press, Oxford (2001)

Buckland, S.T., Baillie, S.R., Dick, J.M., Elston, D.A., Magurran, A.E., Scott, E.M., Smith, R.I., Somerfield, P.J., et al.: How should regional biodiversity be monitored? Environ. Ecol. Stat. **19**(4), 601–626 (2012). doi:10.1007/s10651-012-0202-7

Franklin, J.: Species distribution models in conservation biogeography: developments and challenges. Divers. Distrib. **19**(10), 1217–1223 (2013)

Garrard, G.E., Bekessy, S.A., McCarthy, M.A., Wintle, B.A.: When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. Aust. Ecol. **33**(8), 986–998 (2008)

Guillera-Arroita, G.: Impact of sampling with replacement in occupancy studies with spatial replication. Methods Ecol. Evol. **2**(4), 401–406 (2011)

Guillera-Arroita, G., Lahoz-Monfort, J.J.: Designing studies to detect differences in species occupancy: power analysis under imperfect detection. Methods Ecol. Evol. **3**(5), 860–869 (2012)

Guillera-Arroita, G., Lahoz-Monfort, J.J., Milner-Gulland, E.J., Young, R.P., Nicholson, E.: Using occupancy as a state variable for monitoring the critically endangered Alaotran gentle lemur *Hapalemur alaotrensis*. Endanger. Species Res. **11**, 157–166 (2010a)

Guillera-Arroita, G., Ridout, M.S., Morgan, B.J.T.: Design of occupancy studies with imperfect detection. Methods Ecol. Evol. **1**(2), 131–139 (2010b)

Guillera-Arroita, G., Morgan, B.J.T., Ridout, M.S., Linkie, M.: Species occupancy modeling for detection data collected along a transect. J. Agric. Biol. Environ. Stat. **16**(3), 301–317 (2011)

Guillera-Arroita, G., Hauser, C.E., McCarthy, M.A.: Optimal surveillance strategy for invasive species management when surveys stop after detection. Ecol. Evol. **4**(10), 1751–1760 (2014a)

Guillera-Arroita, G., Lahoz-Monfort, J.J., MacKenzie, D.I., Wintle, B.A., McCarthy, M.A.: Ignoring imperfect detection in biological surveys is dangerous: a response to 'fitting and interpreting occupancy models'. PLoS One **9**(7), e99571 (2014b)

Guillera-Arroita, G., Ridout, M.S., Morgan, B.J.T.: Two-stage Bayesian study design for species occupancy estimation. J. Agric. Biol. Environ. Stat. **19**(2), 278–291 (2014c)

Kendall, W.L., White, G.C.: A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. J. Appl. Ecol. **46**(6), 1182–1188 (2009)

Kéry, M.: Inferring the absence of a species: a case study of snakes. J. Wildl. Manag. **66**(2), 330–338 (2002)

Kéry, M., Schmidt, B.R.: Imperfect detection and its consequences for monitoring for conservation. Community Ecol. **9**(2), 207–216 (2008)

King, R.: Statistical ecology. Annu. Rev. Stat. Appl. **1**, 401–426 (2014)

Lahoz-Monfort, J.J., Guillera-Arroita, G., Wintle, B.A.: Imperfect detection impacts the performance of species distribution models. Glob. Ecol. Biogeogr. **23**, 504–515 (2014)

Lebreton, J.D., Burnham, K.P., Clobert, J., Anderson, D.R.: Modeling survival and testing biological hypotheses using marked animals—a unified approach with case-studies. Ecol. Monogr. **62**(1), 67–118 (1992). doi:10.2307/2937171

MacKenzie, D.I., Royle, J.A.: Designing occupancy studies: general advice and allocating survey effort. J. Appl. Ecol. **42**, 1105–1114 (2005)

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.A.: Estimating site occupancy rates when detection probabilities are less than one. Ecology **83**(8), 2248–2255 (2002)

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E.: Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Academic Press, New York (2006)

Morgan, B.J.T., Revell, D.J., Freeman, S.N.: A note on simplifying likelihoods for site occupancy models. Biometrics **63**, 618–621 (2007)

Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**(4), 308–313 (1965). doi:10.1093/comjnl/7.4.308

Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J.: State-space models of individual animal movement. Trends Ecol. Evol. **23**(2), 87–94 (2008). doi:10.1016/j.tree.2007.10.009

R Development Core Team: R: a language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria (2016). http://www.R-project.org/. Accessed 11 Jan 2016

Royle, J.A.: Site occupancy models with heterogeneous detection probabilities. Biometrics **62**, 97–102 (2006)

Schmidt, B.R., Kery, M., Ursenbacher, S., Hyman, O.J., Collins, J.P.: Site occupancymodels in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. Methods Ecol. Evol. **4**(7), 646–653 (2013). doi:10.1111/2041-210x.12052

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K., Possingham, H.P.: Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecol. Appl. **13**(6), 1790–1801 (2003)

Welsh, A.H., Lindenmayer, D.B., Donnelly, C.F.: Fitting and interpreting occupancy models. PLoS One **8**(1), e52015 (2013)

Wibisono, H.T., Linkie, M., Guillera-Arroita, G., Smith, J.A., Sunarto, G., Pusparini, W., Asriadi, B.P.: Population status of a cryptic top predator: an island-wide assessment of tigers in Sumatran rainforests. PLoS One **6**(11), e25931 (2011)

Yoccoz, N.G., Nichols, J.D., Boulinier, T.: Monitoring of biological diversity in space and time. Trends Ecol. Evol. **16**(8), 446–453 (2001)

Zeller, K.A., Nijhawan, S., Salom-Perez, R., Potosme, S.H., Hines, J.E.: Integrating occupancy modeling and interview data for corridor identification: a case study for jaguars in Nicaragua. Biol. Conserv. **144**(2), 892–901 (2011)